IBM® Technical Disclosure Bulletin Vol. 37 No. 10 October 1994

G06Kg/68B

700
Armonk, NY, US

Using Speech to Correct Optical Character Recognition Output

G06Kg/03

p.665-666 - (2)

Described are several methods whereby speech is used to correct Optical Character Recognition (OCR) output. The methods combine the statistics from the OCR process with speech recognition to create a composite model.

In prior art, many difficulties have been outlined for speech recognition tasks [1,2]. The concept described herein includes methods that are designed to eliminate OCR errors during the distinguishing of like-shaped characters, such as 0 and O, 1 and I, j and g, etc. The following illustrates how speech can be used to correct OCR errors:

1. To input (speak) a corrected word, field, or phrase once the user has identified the error, the user may use a mouse or tab key to identify an incorrect word and then speak it: This can be faster than keying in the correction.
2. To input (speak) a correction once the system has identified an error or area of low confidence. The user might say "accept" or "change to...": (Several OCR systems flag areas of low confidence with color or reverse video.)
3. To identify an erroneous line with speech: For example, "change line 12 to ..."
4. To find and correct an error with speech: For example, "change '100 many elephants' to 'too many elephants'". Or the user might say, "phrase should read, 'too many elephants'". The system would use speech to both find the error and correct it. The error is found by matching a phrase or sentence with the text on screen.
5. To combine the OCR output and the speech recognition system results to create the highest probability combined result: OCR algorithms, like speech recognition algorithms, evaluate multiple competing solutions. Combining the probabilities from both algorithms can yield higher accuracy.

OCR engines make errors unlike errors made by speech recognition engines. For example, a speech recognition engine would rarely mis-hear 100 for too. An OCR engine has three possibilities for a word: 100, too and loo. It displays its highest probability choice in this case "100". The user recognizes the error and speaks the correction. The speech recognition engine also has three competing candidates: to, two and too. While the correct word "too" is neither the highest OCR candidate, nor the highest speech candidate, it is selected because it has the best combined probability. In this instance, it is the only candidate on both lists.

The techniques are applicable to any language and can be used to OCR Kana or Kanji characters, for example. They are also applicable when correcting the results of combined OCR and machine translation systems. For example, situations where a document is OCRed and then translated by machine. In this instance, the statistical models from several systems would be combined; machine translation and speech correction is an especially powerful combination.

References

[1] F. Jelinek, R. L. Mercer, L. R. Behl and J. K. Baker, "Perplexity - A Measure of Difficulty of Speech Recognition Tasks," 94th meeting of the Acoustical Society of America (December 15, 1977). Journal of Acoustical Society of America, Vol. 62, Supl. No. 1, s63 (Fall 1977).

[2] F. Jelinek, R. L. Mercer and L. R. Bahl, "Continuous Speech Recognition Statistical Methods," 549-573 in Handbook of Statistics, Vol. 2: Classification, Pattern Recognition and Reduction of Dimensionality, P. R. Krishnaiah and L. N. Kanal, eds., North-Holland Publishing Company (1982).